Subjective assessments of spherical microphone arrays - Paired comparisons of two arrays designed using different microphone models

Vincent Koehl, Mathieu Paquier, Simeon Delikaris-Manias

Subjective assessments of spherical microphone arrays - Paired comparisons of two arrays designed using different microphone models

V. Koehl[a], M. Paquier[a] and S. Delikaris-Manias[b]

[a]Université de Bretagne Occidentale, 6, avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex 3, France
[b]Ecole Nationale d'Ingénieurs de Brest - CERV, 25, rue Claude Chappe, BP 38, 29280 Plouzané, France
vincent.koehl@univ-brest.fr

Microphone arrays are commonly used to capture sound fields. As the number of sensors forming the array increases, the spatial sampling accuracy at high frequencies improves. Numerous prototypes of spherical arrays were developed over the last years. However, much less attention has been paid to the intrinsic performances of the sensors than to their number and arrangement. This study aims at evaluating the relative performances of two rigid spherical microphone arrays of the exact same size differing only in their capsules (pressure sensors). The two recording systems are based on higher order ambisonics and were used to acquire the exact same sound scene. Four short music excerpts were decoded as various types of audio content (mono, stereo and multichannel) and displayed through dedicated loudspeaker setups. The recordings issued by the two arrays were then to be compared by pairs, on a similarity basis and on a preference one, by twelve expert listeners (sound engineering students). The results showed that the perceived differences and preferences depended on the way stimuli were rendered. These assessments were consistent with those obtained from naive listeners in a previous study, although experts perceived significantly better the differences and reported more pronounced preferences.

1 Introduction

Microphone arrays are nowadays widespread and useful for numerous applications such as spatial audio capture, beamforming or room acoustics measurements. To achieve an accurate acqusition, the array has to be made of a large number of sensors. As a result, the focus has been put on the number of sensors needed [1] and to the way they should be arranged [2] in order to raise the spatial aliasing frequency [3]. However, very few studies have focused so far on practical issues about the the sensors specifications for designing a microphone array. As an example, based on simulations, the use of large membrane microphones should increase the frequency range [2]. Nevertheless, no listening tests were carried out to quantify the perceptual benefit brought by the use of larger membranes (and more generally high quality microphones) in microphone arrays.

The aim of this study is to determine whether the use of different microphone models leads to significant differences in subjective assessments. For this purpose, two microphone arrays differing only in the type of sensors were designed and used to perform music recordings. One of the arrays was designed with assumed better quality capsules than the other one. The exact same sound scene was recorded and short music excerpt issued by the two arrays were to be compared by pairs. The subjective assessments were to be made in terms of similarity and preference to check that the two arrays were clearly perceived as different and that these subjective differences result in modifications of the audio quality. As subjective assessments are strongly related to the content of the excerpts used as test material [4], four different music sequences were to be assessed. The comparisons were to be made by expert listeners on both decoded and raw microphone signals, to determine whether the intrinsic differences between capsules are preserved during the encoding/decoding processes.

The perceptual assessments of these recordings might also be dependent on the way they are presented to listeners. As an example, loudspeakers [5] or room equalization [6] are not appreciated the same way in monophonic, stereophonic or multichannel restitution. The recording were then used to produce various stimulus types (mono, stereo, 5.0) that were displayed on corresponding loudspeaker setups in order to achieve the array comparison under different listening conditions.

2 Experimental setup

2.1 Microphone arrays

Two spherical microphone arrays consisting of eight omnidirectional capsules (pressure sensor) each were designed. The two sets of capsules were each fitted in a rigid sphere which radius was 8 cm. Miniature microphones (DPA 4060 as shown in Figure 1(a)) and small membrane ones (Schoeps CCM2 as shown in Figure 1(b)) were under test. The corresponding arrays will respectively be denoted by *A* and *B* for the rest of this paper.

Based on their frequency responses and signal-to-noise ratios, the small membrane microphones were of assumed higher quality than the miniature ones. The objective differences between the two arrays were also evaluated by characterizing the spherical harmonic components [7]. Differences in the capsules responses were revealed but little difference was observed in the calculated spherical harmonic components.

2.2 Recordings

The two arrays were placed one above the other (and considered coincident for in-plane recordings) in a recording studio where four musicians were placed in a semi-circular arrangement. The quartet consisted of a flute, a clarinet, a double bass and an oboe (from left to right as shown in Figure 2). This setup enabled to acquire the exact same music recordings with the two arrays in order to compare them. Music sequences were recorded (bit depth of 16 bits, sample rate of 48 kHz) and four short excerpts (4 to 7 s) were extracted:

- 1-2: Schumann, opus 94, romance no. 3,
- 3-4: Corelli, opus 4, sonata no. 1.

The recordings were decoded [7] as monophonic (omnidirectional component), stereophonic (1^{st} order decoding equivalent to an XY stereo pair with cardioid capsules at 90°) and multichannel stimuli (3^{rd} and 4^{th} order [3]). In addition to that, mono and stereo stimuli were directly produced using the frontal and anti-diametric (used as a stereo pair [8]) capsules of each array, in order to see whether the decoding process may smooth or sharpen the intrinsic differences between the raw signals. Each of the four excerpts was then used to generate 6 different stimuli that are summarized in Table 1.

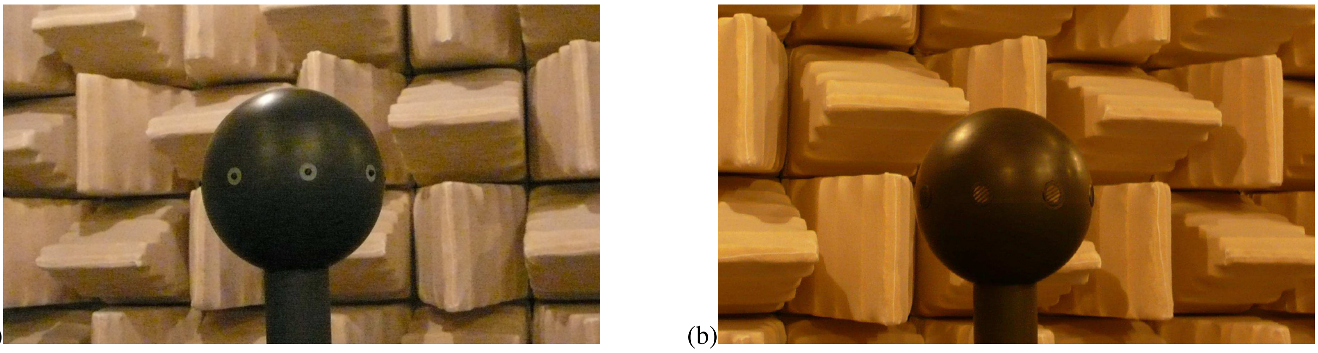

Figure 1: Rigid sphere fitted with miniature (a) and small membrane microphones (b).

Figure 2: Recording setup.

Table 1: Types of stimuli used for array comparison.

Stimulus	Signal type	Signal source
1	monophonic	frontal capsule
2	monophonic	omnidirectional component
3	stereophonic	anti-diametric capsules
4	stereophonic	1^{st} order
5	multichannel	3^{rd} order
6	multichannel	4^{th} order

2.3 Restitution setup

For comparison purpose, the music excerpts were played back in a listening room at a realistic level by using PSI A25M loudspeakers. The recordings issued by the two arrays were to be compared by pairs using the 6 versions (see Table 1) of each excerpt that were restituted on dedicated setups:

- central loudspeaker (*C*) for monophonic signals,
- stereo pair (*L*-*R*) for stereophonic signals,
- ITU setup (*L*-*C*-*R*-*Ls*-*Rs*) for multichannel signals [9],

as shown in Figure 3.

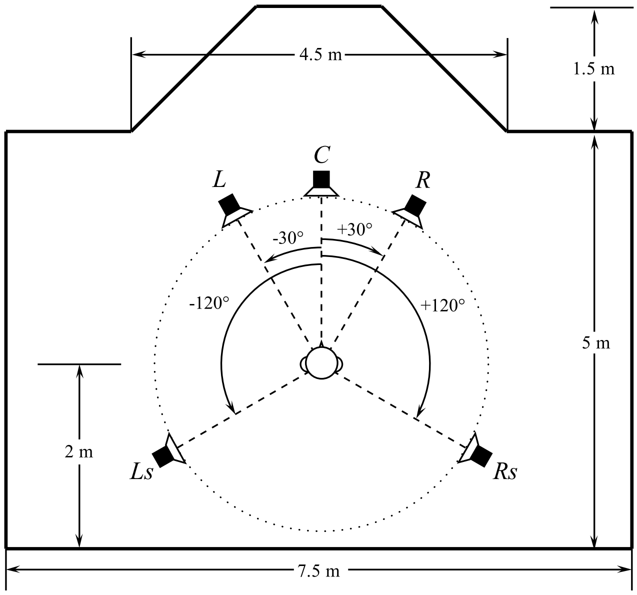

Figure 3: Listening room arrangement for array comparison.

2.4 Listening tests

The excerpts recorded with the two arrays had to be compared by pairs on a similarity basis and on a preference one. Listeners had to compare the stimuli along a continuous scale in both cases. The similarity scale ranged from 0 (identical) to 1 (extremely different) with intermediate labels (slightly different, different, very different). Concerning preference, the answering scale ranged from -1 (*A* strongly preferred) to 1 (*B* strongly preferred) with intermediate labels (*A* preferred, no preference, *B* preferred). The two stimuli were randomized within each pair and were played back through a loudspeaker setup with respect to the stimulus type as indicated in Table 1. Paired-comparisons about similarity and preference occurred then in mono, stereo and multichannel for each music excerpt. The subjects were randomly provided with all the comparisons about one character (similarity or preference) and then about the other one. The test lasted around half an hour and was run using a MATLAB graphical user interface controlled by a touch screen.

2.5 Listeners

Twelve sound engineering students (University of Brest, Master Image & Son), considered as experts [10], took part in this experiment and were remunerated for their participation. The subjects showed normal hearing thresholds based on an audiogram passed in the month preceding this test.

3 Results

3.1 Similarity ratings

An analysis of variance was performed to examine the effects of the stimulus type, the excerpt and their possible interaction on similarity ratings. As shown in Table 2, only the stimulus type proved to have a significant effect which indicates that the differences between the two systems were not equally perceived for all stimulus types.

Table 2: Analysis of variance of similarity ratings by expert listeners.

Source	SS	DF	MS	*F*	*p*
Stimulus type	0.64	5	0.13	2.43	<.05*
Excerpt	0.07	3	0.02	0.46	0.71
S*E	0.59	15	0.04	0.75	0.73
Error	13.92	264	0.05		
Total	13.22	287			

The differences were well perceived for all stimulus types as the average similarity ratings ranged between different (0.5) and very different (0.75), as can be seen in Figure 4. Nevertheless, the differences were not perceived the same way for all stimulus pairs. As an example, in monophonic restitution the differences between the two arrays were significantly better perceived with the omnidirectionnal component (stimulus type 2) than with the raw frontal capsule signal (stimulus type 1).

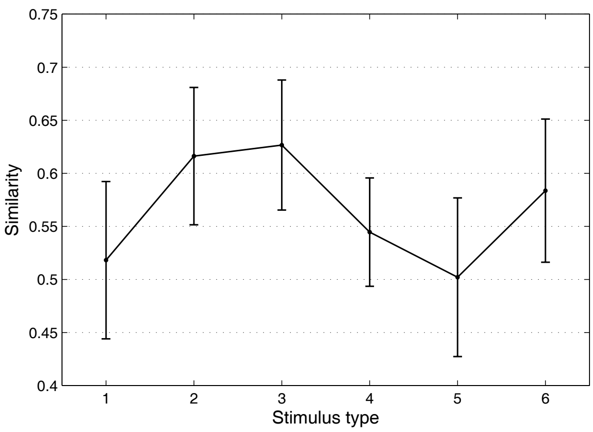

Figure 4: Mean similarity ratings per stimulus type, within their 95% confidence intervals.

3.2 Preference ratings

An analysis of variance was performed to examine the effects of the stimulus type, the excerpt and their possible interaction on preference ratings. Again, and as shown in Table 3, only the stimulus type proved to have a significant effect on listener's preference.

Table 3: Analysis of variance of preference ratings by expert listeners.

Source	SS	DF	MS	*F*	*p*
Stimulus type	22.44	5	4.49	23.56	<.001***
Excerpt	0.58	3	0.19	1.02	0.38
S*E	2.82	15	0.19	0.98	0.47
Error	50.29	264	0.19		
Total	76.13	287			

The significance of this factor effect indicates that the preference of one array over the other depends on the type of stimulus used to achieve the comparison. This effect is mainly due to the fact that the array A is surprisingly prefered when only the frontal capsule is used (stimulus type 1 as can be seen in Figure 5). For all other stimulus types, and as expected because of the high quality sensors, array *B* is prefered.

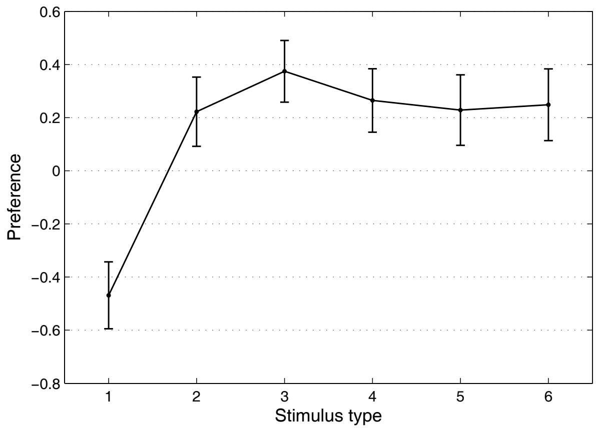

Figure 5: Mean preference ratings per stimulus type, within their 95% confidence intervals.

3.3 Comparison to naive listeners' assessments

These results were compared to subjective evaluations obtained from naive listeners in a previous study [11]. The analyses of variance were carried out again including the listener group (expert or naive) as an experimental factor. As shown in Table 4, the group factor proved to have a significant effect on similarity assessments. As can be seen in Figure 6, the similarity ratings were higher for expert listeners which indicates that the differences between the two arrays were better perceived by this group.

On the other hand, the group factor proved to have no significant effect on preference judgments (Table 5). The mean preference were then the same although the response range was much larger for the expert group (see Figure 7). The non-significance of the group factor over preference ratings is mainly due to the fact that the mean preference is smoothed by the preference towards array *A* occuring only for stimulus type 1 (frontal capsule) and expert listeners. For all other

Table 4: Analysis of variance of similarity ratings by naive and expert listener groups.

Source	SS	DF	MS	F	p
Group	3.46	1	3.46	56.64	<.001***
Stimulus type	1.02	5	0.20	3.11	<.01**
Excerpt	0.33	3	0.11	1.67	0.17
G*S	0.18	5	0.03	0.55	0.74
G*E	0.23	3	0.08	1.18	0.32
S*E	2.04	15	0.14	2.07	<.05*
G*S*E	0.67	15	0.04	0.68	0.8
Error	36.33	552	0.07		
Total	44.34	599			

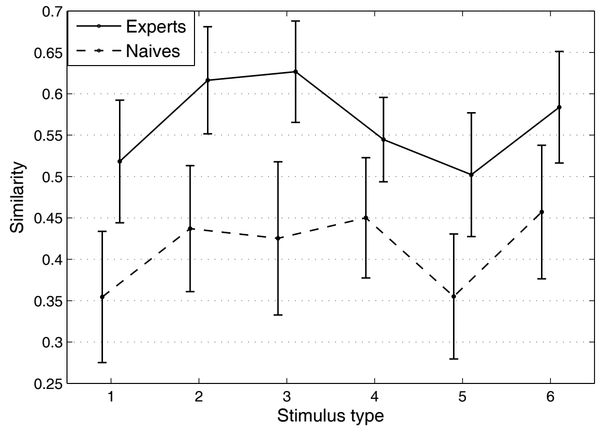

Figure 6: Mean similarity ratings per stimulus type for expert and naive listeners, within their 95% confidence intervals.

stimulus types, the preference is towards array *B*. Nevertheless, the group influence can indirectly be seen through its interaction with the stimulus type. The preference ratings depended then on both the group and the stimulus type. It can be noted here that the preference ratings (towards array *A* or *B*) are significantly more clear-cut for the expert group.

Table 5: Analysis of variance of preference ratings by naive and expert listener groups.

Source	SS	DF	MS	F	p
Group	0.51	1	0.50	2.19	0.14
Stimulus type	15.47	5	3.09	13.36	<.001***
Excerpt	1.53	3	0.51	2.21	0.09
G*S	8.76	5	1.75	7.56	<.001***
G*E	0.38	3	0.13	0.55	0.65
S*E	2.93	15	0.19	0.84	0.63
G*S*E	3.56	15	0.24	1.03	0.43
Error	127.88	552	0.23		
Total	160.22	599			

3.4 Discussion

The experts' assessments about similarity and preference were then in concordance with the naives' ones. Nevertheless, the experts perceived significantly better the difference between the two arrays and attributed a higher mark to the preferred one. The experts similarity and preference judgments were the same for all four music excerpts under test.

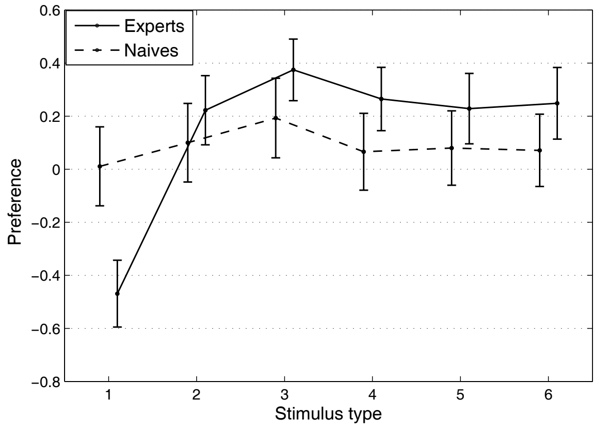

Figure 7: Mean preference ratings per stimulus type for expert and naive listeners, within their 95% confidence intervals.

For all stimulus types, the differences between the two arrays were clearly perceived by the expert listeners. As a rule, the array *B*, which was made of assumed higher quality microphones, was prefered. Nevertheless, the array *A* was surprisingly prefered for monophonic recordings stemming from the frontal capsule, despite its weaker characteristics (frequency response and signal-to-noise ratio [7]). This might be due to the fact that the two microphone arrays are not at the exact same position. The frontal capsule of array *B* might then be placed in an adverse position for this specific recording setup. This drawback doesn't affect the five other stimulus types where multiple capsules contribute to the rendering.

4 Conclusion

The use of different capsules in the two microphone arrays resulted then in perceived differences. For raw and decoded stimuli, in mono, stereo or multichannel restitution, the two arrays were perceived as at least "different" by expert listeners. The differences between the two arrays being well noticed, the audio quality associated to each of them was not the same, based on preference judgments. As a rule, with the exception of monophonic signals issued by the frontal capsule, the array *B*, made of assumed higher quality microphones, was preferred.

Therefore, the use of higher quality capsules in microphone arrays may enhance the perception of sound renderings. This benefit appeared more significant for expert listeners.

Acknowledgments

The authors wish to thank:

- Dr Helmut Wittek (Schoeps Mikrofone GmbH) for the loan of the CCM2 microphones,

- “Les quatre vents” ensemble for permitting to record and use the music excerpt,
- the “Image & Son” department (ISB) of the University of Brest for the loan of the dummy head and the recording studio,
- all the listeners having participated in this experiment.

References

[1] B. Raphaely, “Analysis and design of spherical microphone arrays”, *IEEE Transactions on Speech and Audio Processing* **13**(1), 135–143 (2005)

[2] N. Epain and J. Daniel,“Improving sperical microphone arrays”, *124th Audio Engineering Society convention*, Amsterdam, NL (2008)

[3] S. Moreau, J. Daniel and S. Bertet,“3D sound field recording with higher order ambisonics – Objective measurements and validation of spherical microphone”, *120th Audio Engineering Society convention*, Paris, France (2006)

[4] A. Gabrielsson, U. Rosenberg and H. Sjögren, “Judgments and dimension analyses of perceived sound quality of sound-reproducing systems”, *Journal of the Acoustical Society of America* **55**(4), 854–861 (1974)

[5] F.E. Toole, “Subjective measurements of loudspeakers: A comparison of stereo and mono listening”, *74th Audio Engineering Society convention*, New York, NY, USA (1983)

[6] S. Olive, S. Hess, A. Devantier, “Comparison of loudspeaker-room equalization preferences for multichannel, stereo, and mono reproductions: Are listeners more discriminating in mono?”, *124th Audio Engineering Society convention*, Amsterdam, NL (2008)

[7] S. Delikaris-Manias, V. Koehl, M. Paquier, R. Nicol and J. Daniel, “Does capsule quality matter ? A comparison study between spherical microphone arrays using different types of omnidirectional capsules”, *2nd international symposium on ambisonics and spherical acoustics (Ambisonics Symposium 2010)*, Paris, France (2010)

[8] G. Theile, “On the naturalness of two-channel stereo sound,” *Journal of the Audio Engineering Society* **39**(10), 761–767 (1991)

[9] ITU-R BS.775-1, “Multichannel stereophonic sound system with and without accompanying picture,” *International Telecommunications Union*, Geneva, Switzerland (1994)

[10] ISO 8586-2, “Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Experts”, *International Organization for Standardization*, Geneva, Switzerland (2008)

[11] S. Delikaris-Manias, V. Koehl, M. Paquier, R. Nicol and J. Daniel, “A comparative study of spherical microphone arrays based on subjective assessment of recordings reproduced over different audio systems”, *Forum Acusticum*, Aalborg, Denmark (2011)